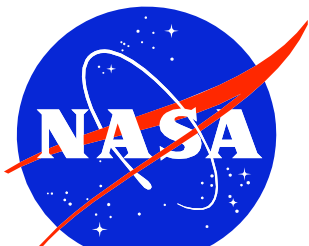


**GAMMA-RAY LARGE AREA
SPACE TELESCOPE
(GLAST)**

**LAT EVENT SUMMARY DATABASE
REQUIREMENTS DOCUMENT**

July 7, 2004



**GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND**

GAMMA-RAY LARGE AREA SPACE TELESCOPE
(GLAST)

LAT EVENT SUMMARY DATABASE
REQUIREMENTS DOCUMENT

July 7, 2004

NASA Goddard Space Flight Center
Greenbelt, Maryland

LAT EVENT SUMMARY DATABASE REQUIREMENTS DOCUMENT

Prepared by:

Robert Schaefer, GSSC

Date

David Davis, Archive Scientist

Date

Concurrence:

Ken Lehtonen, Ground System Manager

Date

Approval:

Jay Norris, GSSC Manager

Date

CHANGE RECORD PAGE**DOCUMENT TITLE:** LAT Event Summary Database Requirements Document**DOCUMENT DATE:** July 7, 2004

ISSUE	DATE	PAGES AFFECTED	DESCRIPTION
Original	7/7/2004	All	

TABLE OF CONTENTS

1	Purpose	1
2	Acronyms	2
3	Glossary	3
4	APPLICABLE DOCUMENTS	4
5	REQUIREMENTS	5
5.1	Description	5
5.1.1	Event Summary Databases	5
5.2	Event Summary Database Requirements	6
5.2.1	General Database Requirements	6
5.2.2	HEASARC Compatibility Requirements	7
5.2.3	Performance for the photon summary database	7
5.2.4	Performance for the comprehensive event database	9
5.3	Performance Goals	11
5.3.1	Summary Photon Database goals	11
5.3.2	Comprehensive Event Summary database	12

1 PURPOSE

This document defines the requirements and goals for the GLAST LAT event summary databases that will serve GLAST data to the GLAST collaboration and the scientific community in general. These databases will be considered as part of the GLAST LAT science analysis environment.

This document will also make the distinction between two databases containing reconstructed LAT Events. There will be a database with all LAT particle events (called the Comprehensive Summary Event Database), and a separate database with only LAT photons (the Summary Photon Database). The photon database will be the main database for scientific analyses.

This document first provides a list of acronyms (§2) and a glossary of key terms and concepts (§3). The relevant documents from which these requirements are derived are listed in §4. Finally, the actual requirements and goals are presented (§5).

2 ACRONYMS AND ABBREVIATIONS

API	Application Program Interface
DBI	Database Interface
DEC	Declination
DBMS	Database Management System
FTE	Full Time Equivalent
GBM	GLAST Burst Monitor
GI	Guest Investigator
GIOC	GBM IOC
GRB	Gamma-Ray Burst
GSFC	Goddard Space Flight Center
GSSC	GLAST Science Support Center
HEASARC	High Energy Astrophysics Science Archive Research Center
IOC	Instrument Operations Center
ISOC	Instrument Science Operations Center (LAT IOC)
LAT	Large Area Telescope
LHEA	Laboratory for High Energy Astrophysics
NASA	National Aeronautics and Space Administration
ODBC	Object Database Connectivity
OGIP	Office of Guest Investigator Programs
PI	Principal Investigator
RA	Right Ascension
SQL	Structured Query Language
SWG	Science Working Group

3 GLOSSARY

Level 0 data—The raw data from the spacecraft after the telemetry packets have been time-ordered and duplicate packets have been removed. The data streams from the spacecraft and the instruments are separated. The processing that produces Level 0 data is called Level 0 processing.

Level 1 data—Level 0 instrument data, which have been processed into discrete particle events. The event tracks have been reconstructed and parameters from the tracks are extracted for use in scientific analysis. The processing that produces Level 1 data is called Level 1 processing

4 APPLICABLE DOCUMENTS

- Science Data Products ICD
- GLAST SSC Functional Requirements Document, 433-RQMT-0002
- GLAST DPWG (Data Products Working Group) Report, Draft 2/25/02, S. Digel
- HEASARC FITS File Recommendations”
(http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html) By the
HEASARC FITS Working Group
- Science Requirements Document, 433-SRD-0001
- Project Data Management Plan (PDMP), 433-PLAN-0009
- GLAST Science Data Products ICD

5 REQUIREMENTS

5.1 Description

The bulk of the publicly available scientific data from the GLAST LAT instrument will come in the form of Event Summary files generated by the ISOC. In order to make this information easily useable (i.e., searchable) they will be stored in a database. This document specifies the requirements and goals on such a database system. The requirements in section 5.2 are derived from usage statistics of currently available high-energy satellite observatory data archives (Chandra, the Rossi X-ray Timing Explorer, etc.), and represent a minimum acceptable level of service. The point to note is that the GLAST event databases will have a harder job for those performance statistics. Data from other observatories come as well separated, prepackaged, narrow field-of-view observations, whereas GLAST has a 2-steradian field of view and will be seeing photons from a large portion of the sky at any given time. Observations of a given source will have to be culled from all of the photons observed. The Summary Photon Database will provide a mechanism to create spatially filtered photon datasets on the fly from survey data. However, it should still be possible to provide service at a much higher level than past observatories, and so in section 5.3 we set performance goals that will provide a speedier environment to better satisfy users.

5.1.1 Event Summary Databases

This is the generic term for databases that contain particle and photon events. The event summary data is expected to be about 99% charged particles and 1% photons. Since it is expected that the scientific community outside of the GLAST collaboration will be mostly interested in the photons, the photon events will be separated out into their own database.

5.1.1.1 Summary Photon Database

This contains an abbreviated list of parameters (energy, time, direction, spacecraft position, etc. as specified in the GLAST Science Data ICD) for gamma rays detected by the LAT. These summaries will be contained in data delivered by the ISOC to the SSC. It is expected that this database will be the most heavily used database by the external scientific community.

5.1.1.2 Comprehensive summary event database

The Summary Photon Database will be a small subset of the Comprehensive Summary Event Database, which includes all particle event summaries (neutral and charged). Each event here will carry more parameters from the LAT reconstruction than photon events in the photon database to allow for more detailed analyses. Again, the summary files used to fill this database are described in the GLAST Science Data Products ICD as products delivered by the ISOC.

5.2 Event Summary Database Requirements

Several types of requirements are described here: general requirements and those specific to the performance of the separate summary photon and comprehensive summary event databases. The performance requirements for the photon database are much more stringent than for the comprehensive summary event database because the photon database will be searched more frequently. Note that this does not preclude having a single database to serve both needs if architecture could be found to meet the performance requirements of both databases simultaneously.

5.2.1 General Database Requirements

The following requirements are independent of the database architecture (including whether or not a Database Management System is used).

5.2.1.1 Searchability Requirements

This section lists the requirements on the types of searches that the database must be able to perform.

5.2.1.1.1 Field Types.

The database must be searchable on fields that are integers, reals, booleans, dates, and times.

5.2.1.1.2 Time Precision

Times must be searchable to microsecond precision.

5.2.1.1.3 Spherical Searches.

Must be searchable by 2-dimensional location on a sphere.

5.2.1.1.4 Quality Field.

Must include data quality as a search criterion.

5.2.1.2 Applications Programming Interface.

There must be an API that allows convenient access to the data in one or more programming languages commonly used in the GLAST LAT project (at least one of the following: C++, JDBC, perl DBI)

5.2.1.3 Remote Access.

The database must be accessible over the internet by computer programs.

5.2.1.4 Mirroring

There must be a procedure to send data updates to more than one copy of the database. These updates must be able to delete, add, and modify data in the database. These updates can be sent to mirrored copies of the database so that all databases can be exactly the same within 6 hours.

5.2.1.5 IEEE Floating point.

It is preferable that real numbers in the database and in any indices use the native IEEE floating-point representation.

5.2.1.6 Upgrade Path

A credible path for upgrading software and/or hardware for the database must be specified. This implies the ability to run the database on standard computers (not specially configured to suit a single DBMS feature) and not writing software which accesses features available in only one single brand of DBMS. The consequence of this for the database is that it must be able to achieve acceptable performance through a relatively standard API. For SQL, this is ANSI standard SQL (currently SQL 99).

5.2.1.7 SQL standards

If an SQL database system is used, the command language must be a dialect of SQL close enough to the ANSI standard so as not to conflict with 5.2.1.6.

5.2.2 HEASARC Compatibility Requirements

5.2.2.1 HEASARC eventual ownership

The database must be able to be turned over to HEASARC staff to run after the mission ends.

5.2.2.1.1 Maintenance Effort

The database must not require excessive amounts of effort (>1.0 FTEs) to maintain and operate.

5.2.2.1.2 Maintenance Fees

The database must not have large (> \$10,000) yearly maintenance fees.

5.2.3 Performance for the Summary Photon Database

5.2.3.1 Service time for public data requests

When the database is queried for photons, the resulting data must be available within a reasonable length of time as defined below.

5.2.3.1.1 Standard Search definition

For purposes of performance specification it is useful to define the standard search, i.e., what we believe will be the most commonly applied database search.

5.2.3.1.1.1 Circular Standard Search

The circular standard search consists of finding all photons appearing to come from within a spherical radius of 15 degrees from some location on the sky sphere within a one-year time window.

5.2.3.1.1.2 Rectangular Standard Search

The rectangular standard search consists of finding all photons coming from a rectangular (in angular coordinates) region whose area is equal to the standard search circle (a 15 degree radius circle is roughly 700 square degrees), within a one-year time window.

5.2.3.1.2 Standard Search Retrieval Time

Standard search requests must retrieve all data in <30 minutes from when searching begins.

5.2.3.1.3 Standard Search Subselection Retrieval Time

A subselection of data from a standard search, where filtering is done on all additional searchable parameters must be served within 45 minutes from the start of the search processing.

5.2.3.1.4 Large Data Request Retrieval Time

Requests for data of size > 2 GBytes (times the number of years in the time window) should be serviced within 3 days. This requirement is meant to allow large requests to be processed during off-peak hours (which will be determined empirically) to prevent bogging down the database.

5.2.3.2 Number of service requests per day.

Must allow for the estimated number of daily requests of data. We make the specification in terms of standard searches.

5.2.3.2.1 Number of standard searches per day

Must be able to do 60 standard search requests in a day.

5.2.3.2.2 Concurrency.

Multiple concurrent search requests (at least 2) must be handled and processed without undue delay. Undue delay is defined as more than 150% of the sum of the time used in processing the same search requests serially.

5.2.3.3 Ingest speed.

After the ingest process starts, data must be available for searching in the database in 10 minutes (per 5 hour's worth of ingested photon summary data).

5.2.3.4 Modification of data in the database

The following processes must be not interrupt access to the database for too long a period (as defined below).

5.2.3.4.1 Ingest Interruptions.

Must allow ingest of new data at the anticipated data rates (about 300,000 events per day) without undue interruption (< 30 minutes) of access to the full database. If the database is broken into subsets, undue delay means either $< 30 / (\text{number of subsets})$ minutes or < 10 minutes, whichever constraint is longer

5.2.3.4.2 Reprocessing Ingest Interruptions.

A reprocessed 5-hour time photon summary file must be able to be inserted into the database without undue interruption. (It must take less than 60 minutes to make the new version of the data available for that interval.)

5.2.3.4.3 Re-indexing time.

A complete re-indexing of the database after a crash must require less than 3 days time per year's worth of data.

5.2.4 Performance for the comprehensive event database

The comprehensive event database, while not the primary database for researchers outside of the GLAST collaboration, still contains much important data useful for scientific analysis and must be accessible to external investigators, but does not need the same high level of performance.

5.2.4.1 Service time for public data requests

When a query is made for data using some search criteria, that data must be available within a reasonable amount of time as defined below. The system must handle at least 1 standard search per day.

5.2.4.1.1 Standard Search Definition

For specifying performance requirements it is useful to define a standard search, i.e., the most common expected search.

5.2.4.1.1.1 Standard Circular Search

The standard circular search consists of finding all particles coming from within a spherical radius of 15 degrees from some location on the sky sphere within a one-year time window.

5.2.4.1.1.2 Standard Rectangular Search

The standard rectangular search consists of finding all particles coming from a rectangle in angular coordinates of area equal to the standard search circle (700 square degrees) within a one-year time window.

5.2.4.1.2 Standard Search Retrieval Time

Requests for data from standard searches must retrieve all data in < 10 hours time.

5.2.4.1.3 Filtered Standard Search Retrieval Time

Requests for standard searches where filtering is done on all parameters will be served within 15 hours.

5.2.4.1.4 Large Data Request Retrieval Time

Requests for data > 20 Gigabytes should be serviced within 7 days. Requests for large amounts of data will require special treatment to avoid bogging down the database.

5.2.4.2 Number of service requests per day.

5.2.4.2.1 Number of standard search requests per day

Must allow for at least 1 request of < 3 Gigabytes of data per day.

5.2.4.2.2 Number of filtered standard search requests.

Must allow for 1 requests of < 3 GB of data using selection criteria beyond a standard search.

5.2.4.2.3 Concurrency

Multiple concurrent searches (at least 2) must be handled and processed without undue delay (< 200% of the time required to do the same requests serially).

5.2.4.3 Ingest speed.

Data must be available for searching in the database after 100 minutes (per 5 hours worth of ingested event summary data). The event summary data file is the standard Level 1 product delivered by the LAT IOC to the SSC.

5.2.4.4 Modifying data in the database

The following processes must be able to be carried out without undue interruption to the access to the full database.

5.2.4.4.1 Ingest interruptions

Must allow ingest of new data at the anticipated data rates (about 3 million events per day)

5.2.4.4.2 Reprocessed data ingest interruptions.

Reprocessing an existing 5 hour time interval of data must be done without undue interruption (< 10 hours)

5.2.4.4.3 Re-indexing time

A complete re-indexing of the database after a rebuild must require less than 1 week's time per years worth of data.

5.3 Performance Goals

5.3.1 Summary Photon Database goals

The following table summarizes performance requirements from 5.2 and shows the performance goals for a more responsive database. The main intent is to support a database that can be searchable interactively.

Action /Requirement	Requirement	Number of Photons	Data size (expected)	Rate (Req.)	Rate (Goal)
Number of standard searches/day	5.2.3.2.1			60/day	1000/day
Standard Searches	5.2.3.1.2	~ 0.15 M	~ 0.3 GB	< 30 min	< 60 sec
Standard Search + filtering on all parameters	5.2.3.1.3	~ 0.15 M	~ 0.3 GB	< 45 min	< 90 sec
Large Data sets (> 10% of DB)	5.2.3.1.4	> 10 M/year	> 2 GB/year	< 3 days/ data year	<3 hours / data year
Number of concurrent search requests handled	5.2.3.2.2			2	20

Ingest new data	5.2.3.4.1	0.05 M	10 M	≤10 min.	≤ 2 min.
Ingest reprocessed data	5.2.3.4.2	0.05 M	10 M	< 60 min.	< 12 min.
Rebuild from scratch	5.2.3.4.3			<3-days/ year data	3 hours/ year data

5.3.2 Comprehensive Event Summary database

Since this database will be used far less frequently than the Summary Photon Database, there is no compelling reason to make it interactively searchable. The goals, however, represent more than an order of magnitude improvement over our minimum requirements

Action	Requirement	Events	Data size	Rate (Req.)	Rate (Goal)
Number of standard searches/day	5.2.4.2.1			1/day	10/day
Standard Searches	5.2.4.1.2	~ 15 M	~ 30 GB	< 10 hours	1 hour
Standard Search + filtering on all parameters	5.2.4.1.3	~ 15 M	~30 GB	< 15 hours	1.5 hours
Large Data sets (> 10% of DB)	5.2.4.1.4	> 1 G/year	> 200 GB/year	< 7 days/ data year	1 day/ data year
Number of concurrent search requests	5.2.4.2.3			2	20
Ingest new data	5.2.4.4.1	5 M	1 GB	≤100 min.	≤ 20 min.
Ingest reprocessed data	5.2.4.4.2	5 M	1 G	<10 hours.	2 hours
Rebuild from scratch	5.2.4.4.3			<7 days/ data year	< 3 days/ data year